

Synthesis and recognition by DNA polymerases of a reactive nucleoside, 1-(2-deoxy- β -D-erythro-pentofuranosyl)-imidazole-4-hydrazide

Heike Strobel, Laurence Dugué, Philippe Marlière¹ and Sylvie Pochet*

Institut Pasteur, Unité de Chimie Organique URA CNRS 2128, 28 rue du docteur Roux, 75724 Paris cedex 15, France and ¹Evologic SA, 89 rue Henri Rochefort, 91000 Evry, France

Received February 14, 2002; Revised and Accepted March 14, 2002

ABSTRACT

We report the synthesis of a new nucleoside, 1-(2-deoxy- β -D-erythro-pentofuranosyl)-imidazole-4-hydrazide (dY^{NH_2}) as a reactive monomer for DNA diversification. The 5'-triphosphate derivative ($dY^{NH_2}TP$, **1**) was evaluated *in vitro* as a substrate for several DNA polymerases. Primer extension reactions showed that $dY^{NH_2}TP$ was well tolerated by KF (exo⁻) and Vent (exo⁻) DNA polymerases. One $dY^{NH_2}MP$ was incorporated opposite each canonical base with an efficiency depending on the template base ($A \approx T > G > C$). Significant elongation after Y^{NH_2} incorporation was observed independently of the $Y^{NH_2}:N$ base pair formed. When the nucleobase Y^{NH_2} was incorporated into synthetic oligodeoxynucleotides via the phosphoramidite derivative **11**, it directed the insertion of natural bases as well as itself. The mutagenicity of $dY^{NH_2}TP$ was evaluated by PCR amplification using Vent (exo⁻) DNA polymerase. The triphosphate $dY^{NH_2}TP$ was preferentially incorporated as a dATP or dGTP analogue and led to misincorporations at frequencies of $\sim 2 \times 10^{-2}$ per base per amplification. A high proportion of transversions with a large distribution of all possible mutations was obtained. The reactivity of the nucleobase Y^{NH_2} within a template with several aldehydes was demonstrated.

INTRODUCTION

In vitro directed evolution of proteins can be achieved by techniques introducing randomised mutations into specific DNA fragments coding for proteins. The goal is to generate a mutant library containing the largest pool of point mutations. By varying the concentrations of Mn^{2+} and dNTPs, the mutation frequency during PCR could be controlled (up to 2%), but a significant bias for transitions over transversions was observed, A and T being mutated more frequently than C and G (1,2). A large number of nucleoside analogues have been proposed as universal bases (reviewed in 3), but they are not always substrates for DNA polymerisation and very few can be

used as nucleoside triphosphates in PCR amplification. Most of the mutagenic nucleotides lead to transitions (4).

We have previously designed a nucleobase with imidazole moiety substituted at position 4 with a carboxamide group (dY ; Fig. 1) which we predicted to act as an ambiguous base (5). Primer elongation reactions catalysed by KF (exo⁻) and *Taq* DNA polymerase showed that this nucleobase directed the incorporation of canonical bases as well as itself, and was also incorporated opposite each of the four canonical bases as a nucleoside triphosphate (6,7). The mutagenic properties of dY as a 5'-triphosphate derivative ($dYTP$) was demonstrated during PCR amplification (8). The introduction of an alkyl chain (methyl or propyl) to the carboxamide extremity yielded monomers (dY^{Me} or dY^{Pr} , respectively) able to be incorporated and copied by DNA polymerases as efficiently as $dYTP$.

The anchoring of variable side motifs to this simplified purine, imidazole-4-carboxamide (Y), was further explored. The introduction of an amino group to the carboxamide extremity yields the nucleobase 4-imidazole hydrazide (Y^{NH_2}) (Fig. 1). By rotation around the hydrazide moiety and glycosidic bond, the nucleobase Y^{NH_2} should pair with all the natural bases as well as with itself according to a base pairing scheme similar to that proposed for Y . Moreover, condensation between the hydrazino function and any aldehyde or ketone should permit to obtain in a single step a large variety of new nucleobases. Their ambiguous pairing properties or their ability to inhibit replication can be evaluated in primer extension reactions.

Here we describe a synthetic access to the 5'-O-triphosphate of 1-(2-deoxy- β -D-erythro-pentofuranosyl)-imidazole-4-hydrazide (**1**) as well as to the phosphoramidite derivative **11**. The ability of $dY^{NH_2}TP$ to be incorporated into DNA was evaluated by primer extension reactions catalysed by commercially available DNA polymerases belonging to the A and B families. We also examined the polymerase recognition of Y^{NH_2} placed in the template strand. The potential for mutagenicity was then evaluated by PCR amplification using $dY^{NH_2}TP$ in place of each of the four canonical bases. The possible reactivity of the nucleobase into DNA was investigated using several aldehydes.

*To whom correspondence should be addressed. Tel: +33 140 61 33 28; Fax: +33 145 68 84 04; Email: spochet@pasteur.fr

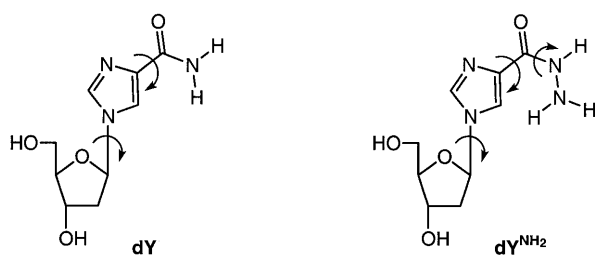


Figure 1. Chemical structure of the simplified purine nucleosides dY and dY^{NH₂}.

MATERIALS AND METHODS

General methods

¹H, ¹³C and ³¹P NMR spectra were recorded on a Bruker AC-300 instrument. Chemical shifts are given in p.p.m. (δ) relative to residual solvent peak in the case of DMSO-d₆ or relative to TMS in the case of CDCl₃ for ¹H and ¹³C, and to 85% phosphoric acid as external standard for ³¹P. Electrospray mass spectra were recorded on a ES/MS platform (VG Biotech-Micromass) working in the negative mode with 50/50 CH₃CN/0.4% NH₄OH or 0.5% Et₃N in water (pH 10). Oligonucleotides were synthesised on an Expedite Millipore DNA synthesiser using standard β-cyanoethyl phosphoramidite chemistry on the 1 μmol scale. The oligonucleotides were separated from failure sequences by preparative HPLC on a Perkin Elmer system with a reverse phase column (C18) using a flow rate of 5.5 ml/min and a linear gradient of CH₃CN (A) in 10 mM triethylammonium acetate buffer (B) at pH 7.5 over 20 min. Reagents and enzymes were purchased from the following sources: the exonuclease-deficient Klenow fragment of *Escherichia coli* DNA polymerase I, T7 DNA polymerase (exo⁺), Vent (exo⁻), T4 polynucleotide kinase and pyrophosphatase from New England Biolabs; *Taq*, *Tli* and *Tth* DNA polymerases from Promega; Pwo from Roche Diagnostics GmbH; Pfu from Stratagene; T7 Sequenase (exo⁻) from Amersham; T4 DNA polymerase from Gibco; Ultra pure dNTPs from Amersham Pharmacia Biotech; and QIAquick PCR purification Kit from Qiagen.

Synthesis of the triphosphate and the phosphoramidite derivatives of the nucleobase Y^{NH₂}

Ethyl 1-[2-deoxy-5-O-(4,4'-dimethoxytrityl)-β-D-erythro-pentofuranosyl]-imidazole-4-carboxylate (3). Compound 2 [ethyl 1-(2-deoxy-β-D-erythro-pentofuranosyl)-imidazole-4-carboxylate] (0.53 g, 2.07 mmol) (5) in dry pyridine (15 ml) was treated with 4,4'-dimethoxytrityl chloride (0.78 g, 2.30 mmol). After stirring for 3 h at room temperature, the mixture was diluted with dichloromethane, washed in turn with aqueous sodium bicarbonate then water, and the organic layer dried (Na₂SO₄). The product was then purified by silica gel chromatography (dichloromethane/methanol). The resulting foam was dissolved in dichloromethane and then petroleum ether was added vigorously. The precipitated product was filtered off and dried *in vacuo* to give compound 3 as a white powder (0.84 g, 73%). ¹H NMR (CDCl₃) δ: 1.30 (t, 3H, CH₃), 2.45 (m, 2H, H2' and H2''), 3.27 (dd, 1H, H5', J = 4.6 Hz, J = 10.2 Hz); 3.37 (dd, 1H, H5''), 3.78 (s, 6H, OCH₃), 4.12 (dd, 1H, H4'), 4.30 (q, 2H, CH₂), 4.55 (m, 1H, H3'), 6.02 (t, 1H, H1', J = 6.3 Hz), 6.82

(d, 4H, H Arom. of DMT), 7.25–7.30 (m, 7H, H Arom. of DMT), 7.40 (m, 2H, H Arom. of DMT), 7.64 and 7.74 (each d, 2H, H2 and H5, J = 1.2 Hz). ¹³C NMR (CDCl₃) δ: 14.16 (CH₃), 41.57 (C2'), 55.04 (OCH₃ of DMT), 60.36 (CH₂), 63.60 (C5'), 72.12 (C3'), 86.13 and 86.17 (C4' and C1'), 86.47 (Cq of DMT), 113.06 (C3 and C5 of DMT), 122.59 (C5 of Im), 126.78 (C4' of DMT), 127.58–129.89 (C Arom. of DMT), 134.07 (C4 of Im), 135.33 and 135.40 (C1 of DMT), 136.21 (C2 of Im), 144.24 (C1' of DMT), 158.39 (C4 of DMT), 162.46 (COOEt). MS (ES) m/z : 557.1 (M-H)⁻.

1-[2-deoxy-5-O-(4,4'-dimethoxytrityl)-β-D-erythro-pento-furanosyl]-imidazole-4-hydrazide (4). Hydrazine hydrate (40 ml) was added to compound 3 (1.0 g, 1.79 mmol) in ethanol (4 ml). The solution was heated at 60°C until completion of the reaction (6 h). Dichloromethane was added to the mixture, the organic layer was washed with water (twice), dried (Na₂SO₄) and evaporated. Compound 4 (0.96 g, 96%) was used in the next step without further purification. MS (ES) m/z : 543.1 (M-H)⁻.

N-Benzyloxycarbonyl-1-[2-deoxy-5-O-(4,4'-dimethoxytrityl)-β-D-erythro-pentofuranosyl]-imidazole-4-hydrazide (5). To a solution of compound 4 (0.96 g, 1.76 mmol) in dichloromethane (40 ml), triethylamine (0.30 ml, 2.15 mmol) and N-benzyloxycarbonyl succinimide (0.62 g, 2.48 mmol) were added. After stirring for 6 h at room temperature, the mixture was diluted with dichloromethane, and washed in turn with aqueous sodium bicarbonate and water. The organic layer was dried (Na₂SO₄) and evaporated. Purification by silica gel chromatography (dichloromethane/methanol), followed by precipitation with petroleum ether gave compound 5 as a white powder (0.94 g, 81%). ¹H NMR (CDCl₃) δ: 2.30–2.50 (m, 2H, H2' and H2''), 3.25 (dd, 1H, H5'), 3.35 (dd, 1H, H5''), 3.78 (s, 6H, OCH₃), 4.08 (m, 1H, H4'), 4.45 (m, 1H, H3'), 5.18 (s, 2H, CH₂Ph), 5.97 (t, 1H, H1', J = 6.3 Hz), 6.83 (d, 5H, H Arom. of DMT and NH), 7.20–7.40 (m, 14H, H Arom. of DMT and Ph), 7.57 and 7.67 (each d, 2H, H2 and H5, J = 1.2 Hz), 8.75 (m, 1H, NH). ¹³C NMR (CDCl₃) δ: 41.43 (C2'), 55.24 (OCH₃ of DMT), 63.74 (C5'), 67.76 (CH₂), 72.22 (C3'), 86.08 and 86.23 (C4' and C1'), 86.63 (Cq of DMT), 113.26 (C3 and C5 of DMT), 120.70 (C5 of Im), 126.99–129.98 (C Arom. of DMT and Ph), 135.17 and 135.53 (C4 of Im and C1 of DMT), 135.71 (C2 of Im), 144.40 (C1' of DMT), 156.27 (CONH), 158.58 (C4 of DMT). MS (ES) m/z : 677.3 (M-H)⁻.

N-Acetyl-N'-benzyloxycarbonyl-1-(2-deoxy-3-O-acetyl-β-D-erythro-pentofuranosyl)-imidazole-4-hydrazide (6). Compound 5 (1.50 g, 2.22 mmol) in pyridine (14 ml) was treated with acetic anhydride (0.34 ml, 4.8 mmol). After stirring for 1 h at room temperature, methanol was added and solvents were removed under reduced pressure. The crude product was treated with 20 ml of 80% acetic acid for 20 min, and the solution was concentrated *in vacuo* and co-evaporated with toluene. Purification by silica gel chromatography gave compound 6 (1.23 g, 1.62 mmol) (73% yield). ¹H NMR (DMSO-d₆) δ: 2.00 (s, 3H, CH₃), 2.30 (s, 3H, CH₃), 2.35–2.60 (m, 2H, H2' and H2''), 3.50 (m, 2H, H5' and H5''), 3.55 (m, 1H, H4'), 5.00 (s, 2H, CH₂Ph), 5.10 (m, 1H, OH5'), 5.20 (s, 2H, H3'), 6.05 (dd, 1H, H1'), 7.25 (m, 5H, Ph), 7.90 and 7.92 (each s, 2H, H2 and H5), 9.12 (s, 1H, NH), 9.72 (s, 1H, NH).

N-Benzyloxycarbonyl-1-(2-deoxy- β -D-erythro-pentofuranosyl)-imidazole-4-hydrazide 5'-O-phosphate (**7**). To a solution of compound **6** (0.33 g, 0.71 mmol) in pyridine (4 ml), a 1 M solution of 2-cyanoethylphosphate in pyridine (1.43 ml) and 1,3-dicyclohexylcarbodiimide (DCC) (0.91 g, 4.41 mmol) were added. After stirring for 2 days at room temperature, water was added to the mixture. The insoluble material was filtered off 1 h later. The filtrate was evaporated and treated with a solution of 1% sodium methylate in methanol (15 ml). After stirring for 2 h at room temperature, the mixture was neutralised by addition of resin Dowex (50WX8, H⁺ form). The filtrate was concentrated and purified by reverse phase HPLC to give compound **7** as the triethylammonium salt (0.18 g, 52%). ¹H NMR (D₂O) δ : 1.26 (t, H, CH₃), 2.60 (m, 2H, H2' and H2''), 3.18 (q, H, CH₂), 4.05 (m, 2H, H5' and H5''), 4.25 (m, 1H, H4'), 4.65 (m, 1H, H3'), 5.22 (s, 2H, CH₂Ph), 6.27 (t, 1H, H1', J = 6.4 Hz), 7.45 (m, 5H, H Arom.), 8.14 and 8.31 (each s, 2H, H2 and H5). ³¹P NMR (D₂O) δ : 0.83. MS (ES) m/z : 456 (M-H)⁻.

N-Benzyloxycarbonyl-1-(2-deoxy- β -D-erythro-pentofuranosyl)-imidazole-4-hydrazide 5'-O-triphosphate (**9**). A solution of DCC (0.22 g, 1.07 mmol) in *t*-butanol (4 ml) was added dropwise over 3 h to a refluxing solution of compound **7** (0.12 g, 0.26 mmol) in 6 ml of *t*-butanol/water (1/1) and morpholine (0.1 ml, 1.15 mmol). Water was added to the reaction mixture, the insoluble material was removed by filtration and the filtrate was evaporated *in vacuo*. Purification on a C18 column (H₂O/methanol) afforded the morpholidate derivative **8** as the DCU salt (0.18 g). ¹H NMR (D₂O) δ : 1.10–1.90 (DCU), 2.60 (m, 2H, H2' and H2''), 3.43 (t, 4H, CH₂N), 3.77 (t, 4H, CH₂O), 3.94 (m, 2H, H5' and H5''), 4.20 (m, 1H, H4'), 4.64 (m, 1H, H3'), 5.21 (s, 2H, CH₂Ph), 6.23 (t, 1H, H1', J = 6.4 Hz), 7.45 (m, 5H, H Arom.), 7.97 and 8.30 (each s, 2H, H2 and H5). ³¹P NMR (D₂O) δ : 8.20. The morpholidate **8** (0.17 g, 0.27 mmol) was dried by co-evaporations with toluene and a solution of pyrophosphoric acid (0.24 g, 1.39 mmol) in DMF (1 ml) containing tri-*n*-butylamine (0.30 ml, 1.28 mmol) was added. After 2 days, the mixture was diluted in 10 mM triethylammonium bicarbonate (TEAB) solution and purified by chromatography on DEAE cellulose (800 ml, 10 mM to 0.5 M TEAB) to give **9** as 3.5 eq. triethylammonium salt (130 mg, 50%). ¹H NMR (D₂O) δ : 2.50 (m, 1H, H2'), 2.60 (m, 1H, H2''), 4.15 (m, 2H, H5' and H5''), 4.25 (m, 1H, H4'), 4.70 (m, 1H, H3'), 5.20 (s, 2H, CH₂), 6.24 (t, 1H, H1', J = 6.6 Hz), 7.45 (m, 5H, Ph), 8.00 and 8.07 (each s, 2H, H2 and H5). ³¹P NMR (D₂O) δ : -22.67 (t, J = 16.5 Hz), -10.76 (d, J = 16.6 Hz), -10.15 (d, J = 16.4 Hz). MS (ES) m/z : 618 (M-H)⁻.

1-(2-Deoxy- β -D-erythro-pentofuranosyl)-imidazole-4-hydrazide 5'-O-triphosphate (**1**). Compound **9** as the triethylammonium salt (120 mg, 0.12 mmol) in water (5 ml) was hydrogenated in the presence of palladium (Pd) black (130 mg) at 4°C. After 2 h, the catalyst was removed by filtration and the solution was lyophilised. The residue was purified by reverse phase HPLC (0–15% A in B) to give **1** as the triethylammonium salt (48 mg). ¹H NMR (D₂O) δ : 2.55 (m, 1H, H2'), 2.65 (m, 1H, H2''), 4.17 (m, 2H, H5' and H5''), 4.25 (m, 1H, H4'), 4.75 (m, 1H, H3'), 6.24 (t, 1H, H1', J = 6.6 Hz), 8.00 and 8.04 (each s, 2H, H2 and H5). ¹³C NMR (D₂O) δ : 9.02 (CH₃), 41.07 (C2'), 47.45 (CH₂), 66.22 and 66.30 (C5'), 71.93 (C3'), 86.66 and 86.78 (C4'), 87.65

(C1'), 121.72 (C5 of Im), 134.46 (C4 of Im), 138.68 (C2 of Im), 164.50 (CONH). ³¹P NMR (D₂O) δ : -22.58 (t, J = 19.8 Hz), -10.72 (d, J = 20 Hz), -10.14 (d, J = 19.7 Hz). MS (ES) m/z : 481 (M-H)⁻.

N,N-Dimethylamino-ethylidene-1-(5-O-dimethoxytrityl-2-deoxy- β -D-erythro-pentofuranosyl)-imidazole-4-hydrazide (**10**). Compound **4** (0.43 g, 0.78 mmol) in dry DMF (8 ml) was treated with *N,N*-dimethylacetamide-dimethylacetal (0.3 ml, 1.9 mmol). After stirring for 3 h at room temperature, the mixture was concentrated *in vacuo* and co-evaporated with xylene. Purification by silica gel chromatography (CH₂Cl₂/MeOH) gave compound **10** as a white powder (0.42 g, 87%). ¹H NMR (CDCl₃) δ : 2.00 (s, 3H, CH₃), 2.27–2.45 (m, 2H, H2' and H2''), 3.00 (s, 6H, CH₃), 3.23 (dd, 1H, H5', J = 5.1 Hz, J = 10.1 Hz), 3.34 (dd, 1H, H5''); 3.78 (s, 6H, OCH₃), 4.03 (dd, 1H, H4'), 4.01 (m, 1H, H3'), 5.96 (t, 1H, H1', J = 6.1 Hz), 6.81 (d, 4H, H Arom. of DMT), 7.25–7.40 (m, 9H, H Arom. of DMT), 7.55 and 7.61 (each d, 2H, H2 and H5, J = 1.2 Hz), 8.98 (s broad, 1H, NH). ¹³C NMR (CDCl₃) δ : 13.28 (CH₃), 38.56 (CH₃), 41.32 (C2'), 55.25 (CH₃ of DMT), 64.03 (C5'), 72.09 (C3'), 85.76 and 86.03 (C4' and C1'), 86.51 (Cq of DMT), 113.23 (C3 and C5 of DMT), 119.15 (C5 of Im), 126.92 (C4' of DMT), 127.94–130.05 (C Arom. of DMT), 135.39–137.23 (C1 of DMT, C4 of Im, C2 of Im), 144.53 (C1' of DMT), 158.55 and 158.62 (C4 of DMT, CONH), 164.12 (C=N). MS (ES) m/z : 612.1 (M-H)⁻.

N,N-Dimethylamino-ethylidene-1-[5-O-dimethoxytrityl-3-(2-cyanoethyl-diisopropyl-phosphoramidite-2-deoxy- β -D-erythro-pentofuranosyl)]-imidazole-4-hydrazide (**11**). 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (115 μ l, 0.45 mmol) and *N,N*-diisopropylethylamine (260 μ l, 1.52 mmol) were added to compound **10** (155 mg, 0.25 mmol) in anhydrous CH₂Cl₂ (5.3 ml). After stirring for 1 h at room temperature, the mixture was diluted with CH₂Cl₂ and successively washed with aqueous sodium bicarbonate and brine. The organic layer was dried (Na₂SO₄), concentrated and the crude product purified by flash chromatography on silica gel (AcOEt/CH₂Cl₂/TEA:45/45/10) affording compound **11** (155 mg, 88%). ³¹P NMR (CDCl₃) δ : 147.03 and 147.20.

Modified oligonucleotides. Oligonucleotides containing the modified base Y^{NH₂} were synthesised using standard phosphoramidite chemistry on 1 μ mol scale. The coupling time for **11** (0.19 M in acetonitrile) was increased to 10 min. After DNA synthesis (trityl on), deprotection was performed with 33% aqueous ammonia at 55°C overnight. Oligonucleotides were purified by reverse phase HPLC, fractions containing the desired product were treated with 80% acetic acid for 20 min and then submitted to further purification by reverse phase HPLC. Analysis using negative-ion ESI-MS (0.5% TEA in water/acetonitrile) revealed experimental masses that are in agreement with theory. MS (ES) m/z for C₂₁₄H₂₇₄N₇₅O₁₃₅P₂₁ (5'-GCATY^{NH₂}GTCATAGCTGTTTCCTG) (expected 6707.4, found 6708) : 6746.10 (M+K-H), 6730.86 (M+Na-H), 6689.91 (M-17). MS (ES) m/z for C₂₁₅H₂₇₅N₇₄O₁₃₆P₂₁ (5'-ATTGY^{NH₂}GTCATAGCTGTTTCCTG) (expected 6722.4, found 6723) : 6759.06 (M+K-H), 6744.65 (M+Na-H), 6704.76 (M-17).

Primer extension reactions

5'-end-labelling of the primer (5'-CAGGAAACAGCTATGAC-3', 1 μ M) was carried out in 70 mM Tris-HCl pH 7.6, 10 mM MgCl_2 and 5 mM DTT by addition of 10 U of T4 polynucleotide kinase and [γ - ^{32}P]ATP (10 Ci/mmol) in 50 μ l final volume. The mixture was incubated for 30 min at 37°C, then for 15 min at 70°C and finally stored at 4°C. Annealing was realised in 30 μ l final volume by incubating 6 μ l of appropriate templates at 10 μ M, 2 μ l of primer at 10 μ M and 10 μ l of ^{32}P -end-labelled primer at 1 μ M for 15 min at 75°C, followed by slowly cooling over 1 h to room temperature. Time-dependent incorporations were carried out at 37°C for T7 Sequenase, KF (exo⁻), T4 or T7 DNA polymerase and at 50°C for thermostable DNA polymerases. A series of 12 μ l reactions were performed for each DNA polymerase. Final concentrations used for elongation reactions were: primer-template, 100 nM; dNTPs, 10, 100 or 500 μ M or 1 mM; and polymerase 0.003, 0.033 or 0.1 U/ μ l. Typical elongation reactions were initiated by mixing 6 μ l of diluted DNA polymerase in 2 \times standard polymerase buffer and 2 μ l of annealed primer-template (600 nM) with 4 μ l of various concentrations of dNTPs. Aliquots of elongation reactions (3 μ l) were quenched in time intervals of 2, 15 and 30 min by adding an equal volume of loading buffer [0.02% (w/v) bromophenol blue, 0.02% (w/v) xylene cyanol FF, 50% (v/v) formamide, 50% (v/v) water]. Before loading onto a 20% denaturing polyacrylamide gel, samples were heated for 1 min at 75°C. After electrophoresis, gels were visualised by autoradiography.

DNA synthesis catalysed by Vent (exo⁻) DNA polymerase

Hybrids were formed by mixing primer and templates as described above. Each primer extension reaction was started by mixing 12 μ l of diluted Vent (exo⁻) DNA polymerase (0.1 U/ μ l) in 2 \times standard buffer [1 \times ThermoPol reaction buffer: 10 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 20 mM Tris-HCl pH 8.8, 2 mM MgSO_4 , 0.1% Triton X-100] and 4 μ l of annealed primer-template (600 nM) with 8 μ l of dY^{NH₂}TP (3 mM). Aliquots (3 μ l) were taken after incubation at 50°C for 2, 15 and 30 min and quenched with the loading buffer. Further elongation was carried out by adding one of the canonical dNTPs depending on the template (1.5 μ l at 10 mM) to the reaction mixture (15 μ l). After additional incubation at 50°C, aliquots of elongation reactions were taken at time intervals of 2, 15 and 30 min and quenched. Products were separated by PAGE and visualised by autoradiography.

Modification of Y^{NH₂} after its enzymatic incorporation into double stranded DNA

The modification of Y^{NH₂} was performed at room temperature by adding a solution of benzaldehyde in $\text{CH}_3\text{COOH}/\text{MeOH}$ (pH 4) to aliquots of the elongation reactions. Samples were heated for 1 min at 75°C before loading onto a 20% denaturated polyacrylamide gel. After electrophoresis, gels were visualised by autoradiography.

Mutagenesis by PCR amplification

The R67 gene coding for the type II dehydrofolate reductase (DHFR) was initially amplified from the parent plasmid pSUR67 (9) by PCR using primer RB (5'-CCCCATGGAACGAAGTAGCAATGAAGTCAG-3') and primer RF (5'-GCGAATTCTTAGTTGATGCGTTCAAGCGCC-3'). The

PCR product was digested and cloned into pTrc99A (Pharmacia) as a *Nco*I and *Eco*RI fragment to yield pTrcR67 (10). For each PCR amplification, reaction mixtures contained 10 mM KCl, 20 mM Tris-HCl pH 8.8, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 2 mM MgSO_4 , 0.1% Triton X-100, 5 ng pTrcR67, 100 pmol of each primer (RB and RF) and 0.05 U/ μ l Vent (exo⁻) DNA polymerase in a final volume of 100 μ l. Typically, dY^{NH₂}TP was used at 1 mM, one of the four canonical nucleotides at a variable concentration (0, 1.25, 2.5, 5, 10 or 20 μ M) and the other three nucleotides at 200 μ M. The cyclic parameters were: 95°C for 5 min; 50 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 10 min; and 72°C for 10 min. All reactions were hot-started. PCR products were analysed by electrophoresis (1.2% agarose gel). Following the mutagenesis reactions, 10% of the PCR products were re-amplified by PCR using primers RB and RF, 200 μ M equimolar concentrations of the canonical bases and 0.025 U/ μ l of Vent (exo⁻) DNA polymerase. The cyclic parameters were: 95°C for 5 min; five cycles of 95°C for 30 s, 55°C for 30 s; 15 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 1 min; and 72°C for 10 min.

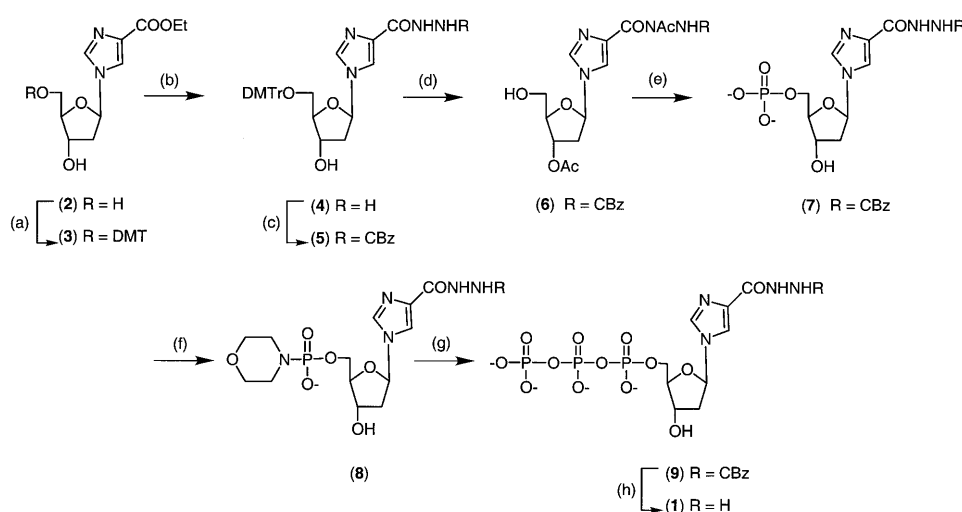
Cloning and DNA sequencing of the amplified fragments

PCR fragments were separated from a 1.2% agarose gel, extracted with QIAquick gel purification kit, digested with *Nco*I and *Eco*RI, purified again with QIAquick and ligated to the expression vector pTrc99A digested with the same enzymes. After transformation (MG1655 cells) and phenotypic expression for 1 h at 37°C, plating an equal volume of the culture on ampicillin (100 μ g/ml) plus trimethoprim (0.3 mM) LB plates yields the functional DHFR variants, while plating on ampicillin alone (100 μ g/ml) yields the functional and defective DHFR variants; both plates contained IPTG (1 mM). After transformation, individual colonies were picked, grown up and sequenced (MWG).

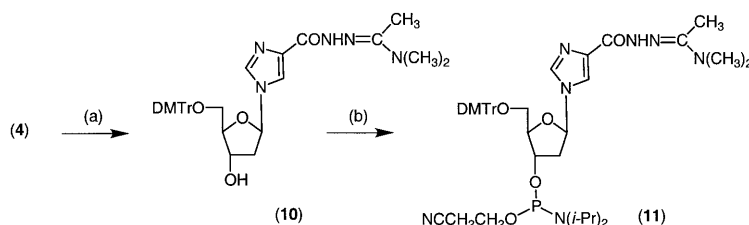
RESULTS AND DISCUSSION

Synthesis of the 5'-triphosphate of imidazole-4-hydrazide deoxyriboside

The synthetic route for 1-(2-deoxy- β -D-erythro-pentofuranosyl)-imidazole-4-hydrazide-5'-O-triphosphate (**1**) is illustrated in Scheme 1. Nucleoside **2** was obtained by an enzymatic transglycosylation using *N*-deoxyribosyltransferase as previously reported (5). The 5'-dimethoxytritylation of **2** in pyridine afforded compound **3** in 73% yield. Treatment of **3** with a large excess of hydrazine hydrate at 60°C yielded compound **4**, which was used in the next step without further purification. The benzyloxycarbonyl group was introduced by reaction of **4** with *N*-benzyloxycarbonyl succinimide affording **5** in 81% yield. Attempts to acetylate selectively the 3'-hydroxyl group were unsuccessful. Acylation with acetic anhydride (1.1 eq.) in acetonitrile in the presence of triethylamine catalysed by 4-dimethylaminopyridine gave a mixture of two major products corresponding to the 3'-O-acetylated compound (24%) and the 3'-O,*N*-diacetylated compound (38%). Complete acetylation was rapidly achieved using acetic anhydride (2.2 eq.) in pyridine. Compound **6** was isolated after detritylation (73% yield from **5**). 5'-Phosphorylation of **6** with 2-cyanoethylphosphate in pyridine in the presence of DCC (11), followed by treatment with 2% sodium methylate in methanol gave the 5'-monophosphate



Scheme 1. Synthesis of **1**. Reagents and conditions: (a) DMT-Cl, pyridine, RT, 3 h; (b) $\text{N}_2\text{H}_4 \cdot \text{H}_2\text{O}$, EtOH, 60°C , 6 h; (c) *N*-CBz-succinimide, Et_3N , CH_2Cl_2 , RT, 6 h; (d) Ac_2O , pyridine, RT, 1 h; (e) 2-cyanoethylphosphate, DCC, pyridine, RT, 2 days, then 2% MeONa in MeOH, RT, 2 h; (f) DCC, morpholine, *t*-BuOH/ H_2O , reflux, 3 h; (g) $\text{H}_3\text{P}_2\text{O}_7$, *n*-Bu $_3\text{N}$, DMF, RT, 2 days; and (h) Pd (black), H_2 , 2 h.



Scheme 2. Synthesis of **11**. Reagents and conditions: (a) $(\text{CH}_3)_2\text{NC}(\text{CH}_3)(\text{OCH}_3)_2$, DMF, RT, 3 h; and (b) CIP(NiPr $_2$)(OCH $_2$ CH $_2$ CN), NEt $_3$, CH_2Cl_2 , RT, 1 h.

7 (52% yield). The triphosphate **9** was obtained in two steps via the intermediate morpholidate **8** (**12**) and isolated after purification by DEAE-cellulose column chromatography with 50% yield. Deprotection of the hydrazino group was achieved by hydrogenolysis on palladium black. Finally, the deprotected triphosphate **1** was purified by reverse phase HPLC and characterized by ^1H , ^{31}P -NMR and mass spectrometries.

Synthesis of oligonucleotides containing the nucleobase Y^{NH_2}

The phosphoramidite derivative **11** was prepared according to Scheme 2. Protection of the hydrazino function of compound **4** by a *N,N*-dimethylacetamidine group afforded compound **10** (87% yield), which was converted to the phosphoramidite **11** by reaction with 2-cyanoethyl-*N,N'*-diisopropylaminochlorophosphoramidite in the presence of *N,N*-diisopropylethylamine in dry dichloromethane (88% yield). The synthesis of oligonucleotides containing this base at predetermined positions was achieved according to standard phosphoramidite chemistry. After reverse phase HPLC purification, the purity of the oligomers was checked by PAGE and electrospray mass spectrometry.

Recognition of the nucleobase Y^{NH_2} by DNA polymerases

The capability of d $\text{Y}^{\text{NH}_2}\text{TP}$ (**1**) to be incorporated into a DNA hybrid was evaluated by primer extension reactions catalysed by DNA polymerases. We tested several commercially available

polymerases. Homopolymeric sequences at the 5' end of the templates were designed to evaluate successive incorporation of the analogue opposite each canonical base. Relative amounts of canonical triphosphates and polymerase were adjusted for each primer extension reaction, in order to facilitate the incorporation of the analogue and minimise the formation of natural base mismatches (non-forcing conditions). Under these conditions, the following DNA polymerases T4, *Pwo*, *Tli*, T7 and T7 Sequenase were not able to incorporate this analogue opposite any natural base. Primer extension reactions catalysed by *Pfu*, *Taq* and *Tth* polymerases resulted in partial incorporation (30%) of d $\text{Y}^{\text{NH}_2}\text{TP}$ opposite T in the template (data not shown) and no significant incorporation was detected opposite the other canonical bases. KF (exo $^-$) and Vent (exo $^-$) polymerases were both able to incorporate d $\text{Y}^{\text{NH}_2}\text{TP}$ opposite all canonical bases. The incorporations of d $\text{Y}^{\text{NH}_2}\text{TP}$ and dYTP catalysed by Vent (exo $^-$) DNA polymerase as a function of time are illustrated in Figure 2. Both analogues were incorporated opposite A or G with a comparable efficiency (Fig. 2A and C). The pattern of incorporation differed in efficiency when these analogues were tested opposite C and T: insertion of d $\text{Y}^{\text{NH}_2}\text{TP}$ was more efficient opposite T than opposite C (Fig. 2B and D). In the conditions tested, there was no successive incorporation of d $\text{Y}^{\text{NH}_2}\text{TP}$, but a band corresponding to primer + 2 opposite T was detected using dYTP. Similar results were observed using KF (exo $^-$).

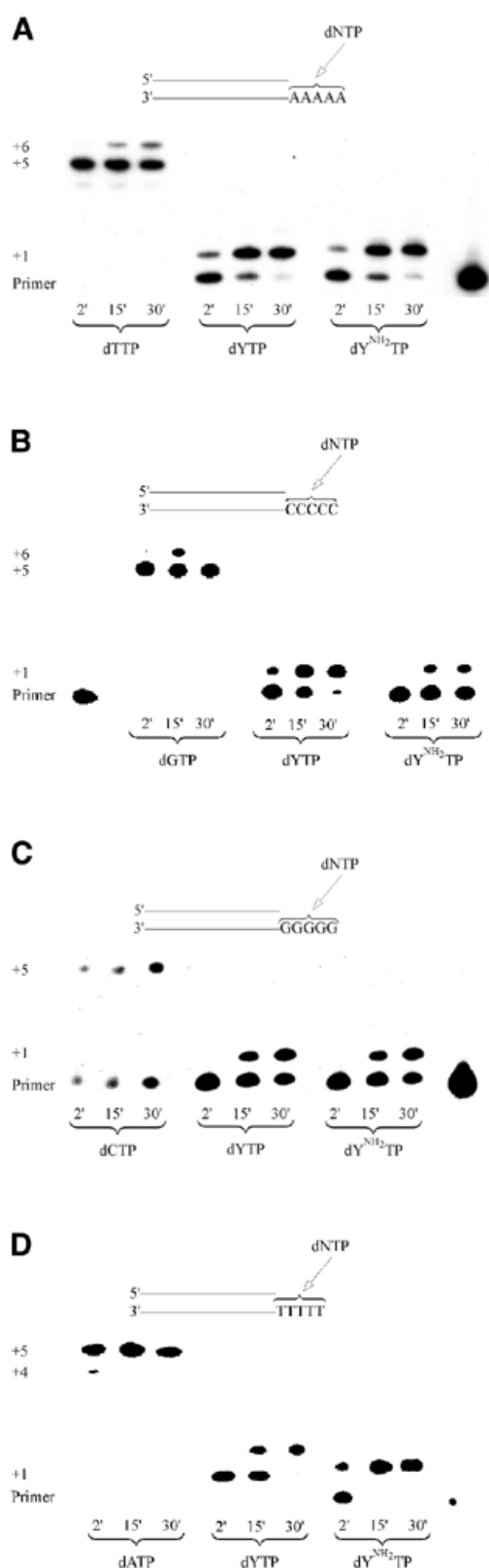


Figure 2. Time-dependent incorporation of dYTP and dY^{NH₂}TP opposite A (A), C (B), G (C) and T (D). Concentrations used were 0.033 U/μl Vent (exo⁻) DNA polymerase, 100 nM primer-template and 10 μM dNTP. The reaction in the absence of dNTPs is realised as control.

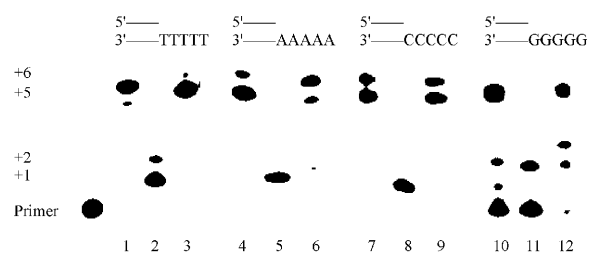


Figure 3. DNA synthesis using canonical or modified nucleoside triphosphates. Conditions: 0.033 U/μl Vent (exo⁻) DNA polymerase, 100 nM primer-template, 100 μM canonical complementary dNTP (lanes 1, 4, 7 and 10) or 1 mM dY^{NH₂}TP (lanes 2, 5, 8 and 11) followed by 100 μM canonical complementary dNTP (lanes 3, 6, 9 and 12), 30 min reaction times. The first lane is the blank reaction in the absence of dNTPs.

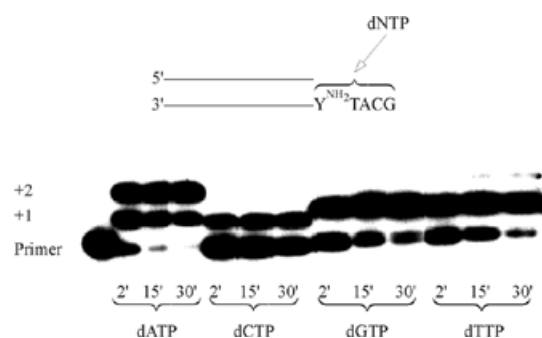


Figure 4. Time-dependent incorporation of canonical nucleoside triphosphates opposite Y^{NH₂} in the template. Conditions: 0.033 U/μl Vent (exo⁻) DNA polymerase, 100 nM primer-template, 100 μM canonical dNTPs, 1 mM dY^{NH₂}TP. Reaction conditions are described in the Materials and Methods. The first lane is the blank reaction in the absence of dNTPs.

We next examined the capacity of Vent (exo⁻) DNA polymerase to further elongate the primer terminated by the analogue Y^{NH₂} with a complementary natural base. Only full-length products were obtained with templates having five A, C or T (Fig. 3, lanes 3, 6 and 9), while a small amount of an intermediate product was detected opposite G (Fig. 3, lane 12). Under these conditions (high triphosphate concentration) limited incorporation of two dY^{NH₂}MP was detected opposite T and G.

These results showed that the triphosphate analogue was incorporated opposite each of the four canonical bases, with an efficiency depending on the nature of the base to be copied (A ≈ T > G > C). Incorporation of two successive dY^{NH₂} residues was observed in the presence of high triphosphate concentration. Further extension with full-length DNA synthesis required the addition of canonical nucleotides.

***In vitro* replication of template Y^{NH₂}**

An oligonucleotide (22mer) containing the nucleobase Y^{NH₂} was used as a template in primer elongation reactions. Figure 4 illustrates the incorporation of canonical bases opposite Y^{NH₂}. dATP, dGTP and dTTP were incorporated opposite Y^{NH₂} resulting in the formation of products at primer + 1 (primer + 2 in the case of dATP) whereas dCTP was less efficiently incorporated. When a mixture of the four canonical nucleotides was used, a full-length product was synthesised. Furthermore, dY^{NH₂}TP was inserted opposite the nucleobase Y^{NH₂} (Fig. 5). No inhibition of replication was observed when dY^{NH₂}TP was

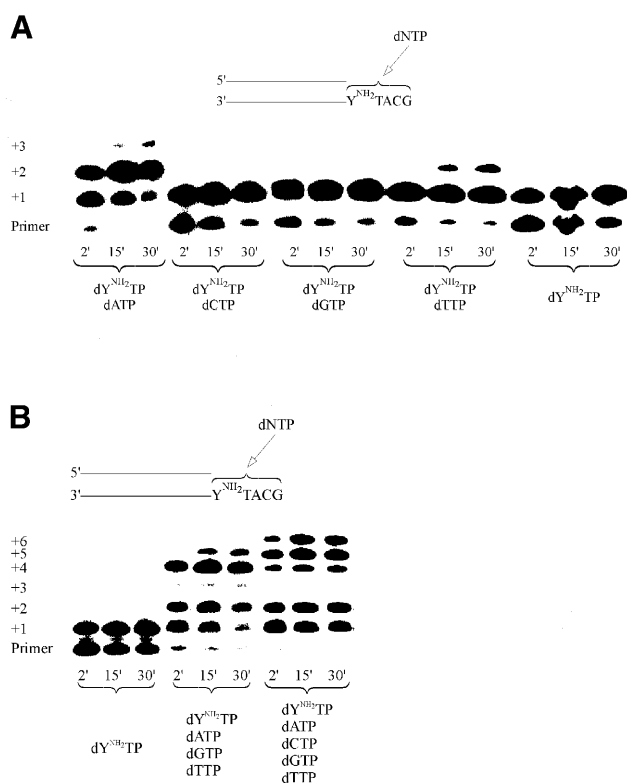


Figure 5. Evaluation of inhibitory effect of dY^{NH₂}TP on the DNA synthesis. Conditions: 0.033 U/μl Vent (exo⁻) DNA polymerase, 100 nM primer-template, 1 mM dY^{NH₂}TP in combination with one canonical dNTP at 100 μM (A) or in combination with three or four canonical dNTPs at 100 μM (B).

used in combination with each of the four canonical bases (Fig. 5A compared with Fig. 4). When a mixture of dY^{NH₂}TP and the four canonical bases were used, full-length DNA synthesis occurred with some pauses depending on the capability of the polymerase to further elongate the Y^{NH₂}:N base pairs formed (Fig. 5B). The products distribution was similar to that obtained with canonical bases. The absence of inhibition is important since the analogue should be used in PCRs in great excess to compete with more efficient incorporation of natural triphosphates. Comparable results were obtained using KF (exo⁻) (data not shown).

Mutagenesis

Target DNA for PCR was the gene coding for the type II DHFR, encoded by the *E.coli* plasmid pTrcR67. No PCR product was detected when the analogue dY^{NH₂}TP totally substituted one of the four dNTPs. In order to force the incorporation of the analogue, its concentration was increased to 1 mM and a low concentration of the substituted dNTP (between 1.25 and 20 μM) was added. A PCR product could only be detected when dATP or dGTP was substituted by dY^{NH₂}TP. Since biased dNTP pools are known to be mutagenic in PCR, amplification without dY^{NH₂}TP was conducted as a control. Mutagenic PCR products were chased in a second round of PCR using only equimolar concentrations of canonical dNTPs. These PCR products were then cloned and screened for DHFR activity. R67 DHFR confers resistance to the antibiotic trimethoprim (trim), unlike its *E.coli* counterpart. Plating out R67 DHFR recombinants on trimethoprim plus ampicillin (amp) plates yields the functional DHFR variants, while

plating on ampicillin alone yields the total collection of variants, functional and defective. Therefore the mutagenic effects of the reaction conditions were reflected in trim^R/amp^R ratios <1 (Table 1).

Two successive PCR amplifications using Vent (exo⁻) and the four canonical nucleotides at the same concentration (200 μM) resulted in a spontaneous mutagenesis (trim^R/amp^R = 0.83) (reaction 1). When dY^{NH₂}TP was added to the four canonical bases, the ratio decreased to 0.65 implying that the analogue competed with the normal bases and induced some mutations. When dATP was present at a concentration of 20, 10 or 5 μM, dY^{NH₂}TP increased the number of mutations (reactions 2–4). Below 5 μM of dATP (reaction 5), no significant mutagenic effect could be assigned to dY^{NH₂}TP. When dY^{NH₂}TP substituted dGTP, an excess of mutations over the control was measured at 20 μM dGTP (reaction 6). No evidence of a significantly altered ratio was found when dGTP was added below 20 μM (reactions 7–9).

Clones obtained from reactions 2 and 6 were sequenced. The resulting mutations are given in Table 2. When dY^{NH₂}TP was used to replace dATP (reaction 2 + dY^{NH₂}TP) the mutation frequency was 2×10^{-2} per base per amplification, while that obtained with biased dATP concentration (reaction 2 control) was 9×10^{-3} per base per amplification. The major substitutions were A→G, T→C (30%), G→A, C→T (28%) and A→T, T→A (20%), with a slight excess of transitions (Ti) (58%) over transversions (Tv) (42%). A similar Ti/Tv ratio was obtained in the control reaction (61 and 39%, respectively), but with another composition. The use of substandard dATP concentration resulted in 35% of all possible A replacements (A→N), while the addition of dY^{NH₂}TP increased the percentage of A replacements (55%), indicating that dY^{NH₂}TP was accepted as an A mimic. The addition of dY^{NH₂}TP to substandard dGTP concentration raised the mutation frequency from 8×10^{-3} (reaction 6 control) to 1.3×10^{-2} (reaction 6 + dY^{NH₂}TP) per base per amplification. The major mutations were G→T, C→A (41%) and G→A, C→T (35%). If we consider all possible G replacements (G→N), a similar percentage was obtained using biased dGTP concentration with or without dY^{NH₂}TP (86% and 90%, respectively). Also, the incorporation of dY^{NH₂}TP as a G mimic could not be directly correlated with the type of mutations. However, the addition of dY^{NH₂}TP increased the total number of transversions (65% compared with 51%) with a higher proportion of A→T, T→A and G→C, C→G mutations. Among the 71 clones sequenced there was no insertion, but a small number of deletions were found. These appeared to be independent of the presence of dY^{NH₂}TP and could be attributed to the biased conditions.

Amino acid sequences of sequenced clones 136–160 (reaction 2 + dY^{NH₂}TP) and 278–301 (reaction 6 + dY^{NH₂}TP) are given in Table 3. These data show that mutagenesis using dY^{NH₂}TP as replacement for dATP or dGTP yielded randomly distributed amino acid changes with no hotspot.

Reactivity of the hydrazine function toward aldehyde

The possible reaction of Y^{NH₂} once it is incorporated into DNA with aromatic aldehydes of different sizes (benzaldehyde, *trans*-cinnamaldehyde and anthracene-9-carboxaldehyde) as well as with a fluorescent compound (fluorescamine) was investigated. Each of the four primer-templates was elongated in the presence of dY^{NH₂}TP in the conditions described

Table 1. Trimethoprim-resistant mutant frequencies resulting from mutagenic PCR with dY^{NH₂}TP

Reaction	dNTP/ μ M				trim ^R :amp ^R		trim ^R :amp ^R		dY ^{NH₂} TP (mM)
	A	C	G	T	PCR control		PCR + dY ^{NH₂} TP		
					Ratio	%	Ratio	%	
1	200	200	200	200	6048:7280	83	2903:4464	65	1
2	20	200	200	200	3034:3832	79	301:742	41	1
3	10	200	200	200	566:907	62	851:2096	41	1
4	5	200	200	200	133:137	97	56:82	68	1
5	2.5	200	200	200	191:231	82	103:136	76	1
6	200	200	20	200	997:1534	65	131:347	38	1
7	200	200	10	200	160:194	82	292:339	86	1
8	200	200	5	200	1254:1336	93	249:350	71	1
9	200	200	2.5	200	672:624	107	568:786	72	1

trim^R and amp^R, trimethoprim- and ampicillin-resistant colonies. The ratio trim^R:amp^R yields the proportion of functional variants following hypermutagenesis.

Table 2. Base substitution frequency for hypermutagenised R67 DHFR

Reaction ^a	Colonies sequenced	Nb. mut. ^b (del)	Mutation frequency ^c	Ti/Tv ^d	Transition Nb (%)			Transversion Nb (%)		
					A \rightarrow G	G \rightarrow A	A \rightarrow T	A \rightarrow C	G \rightarrow C	G \rightarrow T
					T \rightarrow C	C \rightarrow T	T \rightarrow A	T \rightarrow G	C \rightarrow G	C \rightarrow A
2 + dY ^{NH₂} TP	22	105 (1)	2.0×10^{-2}	61/44	32 (30%)	29 (28%)	21 (20%)	5 (5%)	4 (4%)	14 (13%)
2 control	12	26 (1)	0.9×10^{-2}	16/10	6 (23%)	10 (38%)	3 (12%)	0 (0%)	2 (8%)	5 (19%)
6 + dY ^{NH₂} TP	21	63 (6)	1.3×10^{-2}	22/41	0 (0%)	22 (35%)	7 (11%)	2 (3%)	6 (10%)	26 (41%)
6 control	16	31 (4)	0.8×10^{-2}	15/16	1 (3%)	14 (45%)	1 (3%)	1 (3%)	1 (3%)	13 (42%)

^aReactions are those given in Table 1.

^bTotal number of mutations (Ti + Tv) (number of deletions are given in parentheses).

^cMutation frequency is the number of mutations scored divided by the product of the number of clone sequences and the target length (237 bp).

^dNumber of transitions and transversions.

previously. The elongation products were then treated with an excess of aldehyde in alcohol. Gel electrophoresis analysis shows that the coupling reactions were completed within 2 min as shown in Figure 6 when dY^{NH₂}TP was incorporated opposite T. We also checked that the nucleobase Y^{NH₂} located within the primer could also be coupled with benzaldehyde (data not shown).

CONCLUSION

We designed and synthesised 1-(2-deoxy- β -D-*erythro*-pentofuranosyl)-imidazole-4-hydrazide (Y^{NH₂}), which is related to the previously reported ambiguous base, 1-(2-deoxy- β -D-*erythro*-pentofuranosyl)-imidazole-4-carboxamide (Y) (5). This analogue was expected to pair with the canonical bases according to a base-pairing scheme analogous to that of Y. The presence of the hydrazino function allows to condense the nucleobase with various aldehyde or ketone to generate a family of nucleotides.

The effect of the hydrazino function attached to the carboxy-imidazole moiety on the recognition by DNA polymerases was

examined. The new base possesses the characteristics of an ambiguous one, i.e. the ability to replace the canonical bases in DNA replication reactions both as a nucleoside triphosphate and as a template base. The mutagenic effect of the analogue was measured during PCR experiments when dY^{NH₂}TP replaced dATP or dGTP. The resulting substitution frequencies of $1.3\text{--}2 \times 10^{-2}$ per base per amplification are comparable with those obtained with other random mutagenesis procedures. However, the presence of dY^{NH₂}TP raises the proportion of transversions and allows to produce almost all possible mutations.

The specific reactivity of the nucleobase Y^{NH₂} into DNA toward aromatic aldehydes of different size as well as fluorescamine was demonstrated. The presence of the hydrazino function permits the diversification of the imidazole motif by reaction with any aldehyde or ketone. Oligomer libraries could be generated by enzymatic incorporation of the hydrazide triphosphate catalysed by DNA polymerase, or by chemical synthesis using the phosphoramidite derivative, followed by modification of the hydrazide function. Oligomers containing

8. Sala, M., Pezo, V., Pochet, S. and Wain-Hobson, S. (1996) Ambiguous base pairing of the purine analogue 1-(2-deoxy-beta-D-ribofuranosyl)-imidazole-4-carboxamide during PCR. *Nucleic Acids Res.*, **24**, 3302–3306.
9. Pattishall, K.H., Acar, J., Burchall, J.J., Goldstein, F.W. and Harvey, R.J. (1977) Two distinct types of trimethoprim-resistant dihydrofolate reductase specified by R-plasmids of different compatibility groups. *J. Biol. Chem.*, **252**, 2319–2323.
10. Martinez, M.A., Pézo, V., Marlière, P. and Wain-Hobson, S. (1996) exploring the functional robustness of an enzyme by *in vivo* evolution. *EMBO J.*, **15**, 1203–1210.
11. Tener, G.M. (1961) 2-Cyanoethyl phosphate and its use in the synthesis of phosphate esters. *J. Am. Chem. Soc.*, **83**, 159–168.
12. Moffatt, J.G. (1964) A general synthesis of nucleoside-5'-triphosphates. *Can. J. Chem.*, **42**, 599–604.